# 学习、研究和实习汇报表

学号：SA17011009 姓名：陈伟 研究期间：2017.08 ~2017.09

填报说明：1. 月报在每月1号前后3天填报，半月报每月15 日前后2天填报，按规定的样式累计汇报；

2. 研1和在外地的学生填写月报，研2以上的本地学生填写半月报；

3. 有表格之外的内容可以在表后附页；

4. 按时完成汇报，不能敷衍了事。

### 一、工作计划及完成情况

|  |  |
| --- | --- |
| （一）工作计划 | （二）完成情况 |
| 2017.08   1. 调研第二代和第三代基因组测序原理及特点； 2. 搜集第三代基因组测序及序列比对算法文章； 3. 阅读龙广宇师兄的生物医学文献挖掘文章。 | 2017.09   1. 目前运用广泛的测序技术还是基于二代测序，三代测序主要运用在小基因组组装； 2. 研究三代测序mashmap算法，学习MinHash和winnowing算法； 3. 查看，运行mashmap代码，探究运行结果； 4. 搜集了关于生物文献的其他研究热点，如宏基因组。 |

### 二、存在问题和收获

|  |  |
| --- | --- |
| （一）存在问题 | （二）研究收获 |
| 2017.08   1. 人类基因序列数据量很大，比对时间主要花费在验证上，目前很多算法是采用减少验证时间来加快比对； 2. 三代测序仪一直在改进，数据存在不同，有的算法效果好是因为只采用了模型数据。 | 2017.09   1. 能不能在比对过程中直接做下游分析； 2. 是否能改进验证算法。 |

### 三、阅读文献及简评（月报≧3篇，半月报≧2篇）

|  |
| --- |
| 2017.08   1. Jain C, Dilthey A, Koren S, et al. A fast approximate algorithm for mapping long reads to large reference databases[C]//International Conference on Research in Computational Molecular Biology. Springer, Cham, 2017: 66-81.   评：利用poission分布结合两者hash减少待验证序列   1. Ukiyama N，Imai H．Parallel multiple alignments and their implementation on CMS[J]．Genome Informatics 4，1993，103—1O8．Schleimer S, Wilkerson D S, Aiken A. Winnowing: local algorithms for document fingerprinting[C]//Proceedings of the 2003 ACM SIGMOD international conference on Management of data. ACM, 2003: 76-85.   评：winnowing算法原理   1. CRF与词典相结合的疾病命名实体识别。   评：能否用相同方法识别蛋白质类型 |

### 四、投稿和发表论文情况

|  |
| --- |
| （一）发表论文： |
|  |
| （二）投稿论文： |
|  |